

Answering Seven Questions

Aayush Patil(22220)

August 28, 2024

1 Question 1

After running the one level Girvan Newman algorithm(ONE ITERATION) on the wikipote.txt dataset I got 24 communities. The edge removed was (2470,2565), and the modularity of the graph was (0.00050143866).

On running the one level Girvan Newman algorithm on the lastfm-asia-edges.csv dataset the edge removed was (3103,7237) and modularity of the graph was 0.0 and the no. of communities were 1 .

Due to the huge difference in the size of both datasets, the algorithm completed its one iteration on the lastfm dataset very quickly(0.5 sec) as compared to the wikipote dataset(1.2 seconds). The times i have reported are after using multiprocessing on a 24 CPU cores machine. On an average the time taken are more (approximately 6 times of these values)

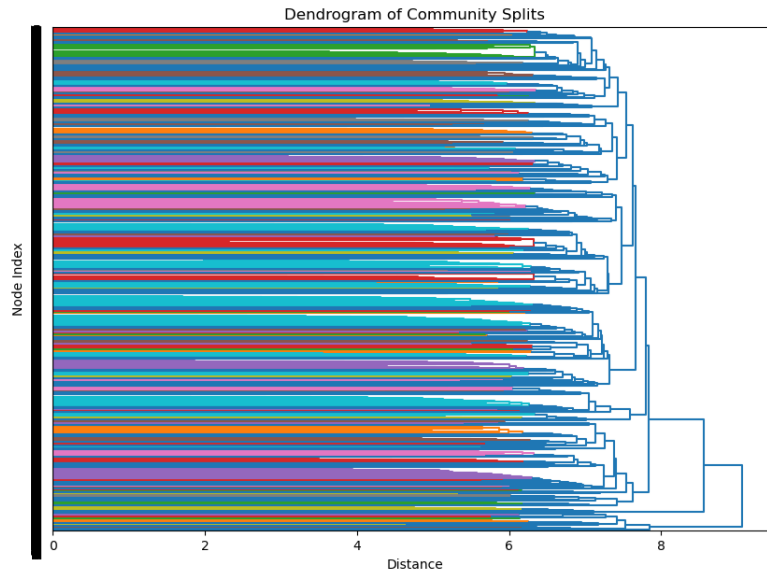
2 Question 2

First i used a modularity based stopping criteria(stop if modularity decreases or increase by an amount less than threshold) in the algorithm, but it took too long to converge, and didn't give meaningful results.

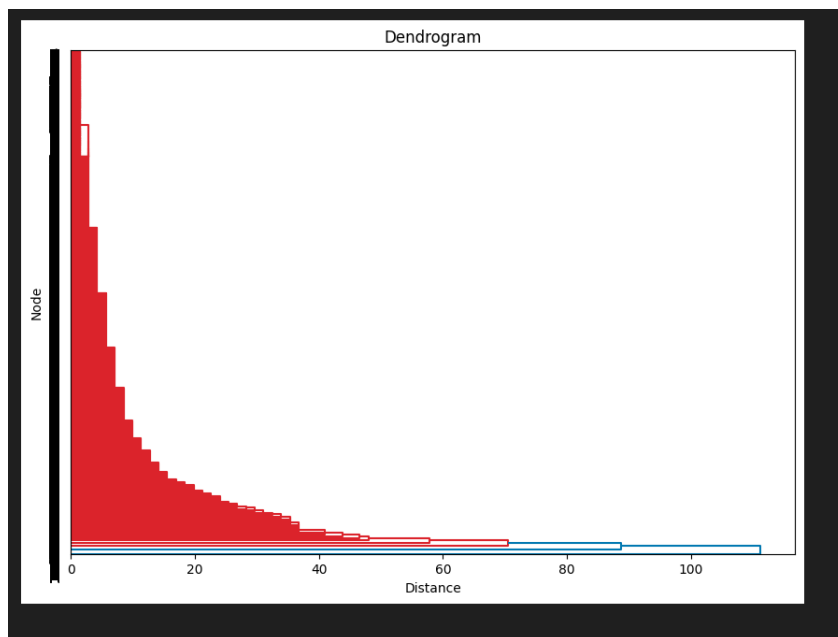
Then I switched to a criteria which stops the algorithm if the number of communities don't change in specific number of consecutive iterations.

3 Question 3

Here you provide the answer to the third question.



This dendrogram is for the LASTFM DATASET



This dendrogram is for the WIKIVOTE dataset

4 Question 4

The louvain algorithm gave 313 communities after one iteration on the wikivote.txt

The louvain algorithm gave 1697 communities after one iteration on the lastfm dataset.

5 Question 5

To pick the best decomposition of nodes into communities using the Girvan-Newman and Louvain algorithms, you should evaluate the quality of the community structures produced by each algorithm. Here's a step-by-step guide for both algorithms:

1. Girvan-Newman Algorithm

The Girvan-Newman algorithm is a divisive method that progressively removes edges with the highest betweenness centrality to detect community structure. Here's how to assess the best community decomposition:

**a. Modularity Score:

- Calculate Modularity: Compute the modularity score for each community structure obtained at different stages of the algorithm. Modularity quantifies the strength of the division of a network into communities.
- Choose the Highest Modularity: The best decomposition is often associated with the highest modularity score.

**b. Number of Communities:

- Stability: Look for a stable number of communities. If the number of communities changes frequently across iterations, the structure might not be stable.

**c. Edge Betweenness:

- Plot Edge Betweenness: Analyze how edge betweenness changes over iterations. This can help understand if the removal process is consistently leading to meaningful community divisions.

**d. Community Size:

- Check Community Sizes: Verify if the community sizes are reasonable. Extremely small or large communities might indicate overfitting or underfitting.

2. Louvain Algorithm

The Louvain algorithm is a hierarchical clustering method that optimizes modularity through iterative refinement. Here's how to pick the best community structure:

****a. Modularity Score:**

- **Compute Modularity:** During the final iteration, compute the modularity score of the final community structure. The higher the modularity, the better the community structure.

****b. Hierarchical Levels:**

- **Analyze Hierarchical Levels:** The Louvain algorithm produces a hierarchy of communities. Evaluate the modularity at different levels to select the most meaningful decomposition.

****c. Stability:**

- **Check Stability Across Runs:** Run the Louvain algorithm multiple times with different initial conditions and compare the results. Consistent results across runs generally indicate a reliable decomposition.

****d. Community Sizes:**

- **Assess Community Sizes:** Similar to Girvan-Newman, verify that the community sizes are appropriate. A balanced distribution is often preferable.

General Evaluation Criteria:

- **Modularity:** It's a primary metric for assessing community structure quality. Both algorithms aim to maximize modularity, so higher scores typically indicate better decompositions.
- **Visual Inspection:** Plot the community structure if possible. Visualizing the network with communities can provide intuitive insights into the quality of the decomposition.

By combining these evaluations, you can pick decomposition of nodes .

6 Question 6

The louvain algorithm was much faster than the girvan newman algorithm. The louvain algorithm almost took 200 minutes for the wiki dataset while it took 7 days for girvan newman. and for the lastfm dataset it took 11 mins for louvain and 220 minutes for the giravn newman algorithm.

7 Question 7

The louvain gave rise to better communities , because it uses modularity calculations rather than edge betweenness which makes it faster for community formation rather than girvan newman. The girvan newman is more accurate for smaller datasets, and more time consuming for larger datasets.